

WebGRMS: Prototype software for web-based mapping of biological collections

Stephanie L. Greene · Toshimi Minoura ·
Jeffrey J. Steiner · Gwen Pentacost

Received: 23 August 2005 / Accepted: 2 May 2006 / Published online: 16 August 2006
© Springer Science+Business Media B.V. 2006

Abstract Biological collections are gaining recognition as priceless sources of information about the historic distribution and diversity of life. The Internet is emerging as the major venue for sharing biodiversity information since it supports globalization and broad-scale interoperability. This research demonstrates how a Web-based mapping application for biological collections was developed using WebGD, an open-source software development tool, and illustrates how simple spatial analysis help collection users describe the range of ecogeographic variation in collections and customize the selection of accessions based on georeferenced variables. Our prototype can be viewed at http://www.yukon.een.orst.edu/~greene/ms_apps/home/index.htm. The demonstration site has three functional areas: (i) Query, (ii) Analyze Collections, and (iii) Add Data. The application was developed relatively quickly and at a low cost, since the complex workings for delivering GIS functions over the Web were an internal part of the WebGD framework. Because it was based on open-source code, costs were greatly decreased compared to commercially available software. In its current form, the prototype WebGRMS application provides users interested in *Medicago* and *Trifolium* germplasm with an innovative method to better understand the germplasm collections. More

S. L. Greene (✉)

National Temperate Forage Legume Germplasm Resources Unit, USDA-Agricultural Research Service, 24106 North Bunn Road, Prosser, WA 99350, USA
e-mail: greenes@wsu.edu

T. Minoura

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA

J. J. Steiner

GWCC-BLTSVL 5601 Sunnyside Avenue, Beltsville, MD, 20705-5140, USA

G. Pentacost

Western Regional Plant Introduction Station, USDA-Agricultural Research Service, Washington State University, Johnson Hall, Room 59, Pullman, WA 99164, USA

importantly, we hope the prototype provides a glimpse into the future of Web-based spatial analysis of biological collections.

Keywords Biodiversity informatics · Geographic information system (GIS) · Germplasm collection · *Medicago* · Plant genetic resources · *Trifolium*

Abbreviations

ARS	United States Agricultural Research Service
GARP	Genetic Algorithm for Rule-Set Production
GBIF	Global Biodiversity Information Facility
GIS	geographic information systems
GRIN	Germplasm Resources Information Network
NPGS	National Plant Germplasm System
PGR	Plant Genetic Resources
TSA	The Species Analyst Project
USDA	United States Department of Agriculture

Introduction

Due to the sheer numbers of samples and the data associated with them, biological collections are gaining recognition as priceless sources of information about species' historic distributions and diversity (Edwards et al. 2000; Krishtalka and Humphrey 2000; Ponder et al. 2001). Botanists have historically used collections of dried specimens held in herbaria, living plants, and seeds to help identify and understand the plants around us. Similarly, plant breeders have used ex situ plant germplasm collections as sources for potentially useful traits to increase agricultural production. Increasingly, ex situ plant germplasm collections are also being recognized as important tools for the conservation and reintroduction of plant species (Greene and Morris 2001; Guerrant et al. 2004).

Because biological collections are geographically dispersed and their documentation widely disparate in digital completeness and database architecture, the Internet is emerging as the major venue for sharing biodiversity information since it supports globalization and broad-scale interoperability (Bisby 2000). Natural history museums are leading the way towards new computational procedures that establish interoperability among widely distributed databases (Graham et al. 2004; Lowe 2004). Although database schemas may vary, nearly all biological collections have common shared sets of biological and ecogeographic information. In the United States, The Biodiversity Research Center Informatics group at the University of Kansas Natural History Museum (<http://specifysoftware.org/Informatics>) has developed collection-specific metadata standards (e.g., various versions of the Darwin Core) and information exchange protocols such as Z39.50. More recently, DiGIR (Distributed Generic Information Retrieval) was developed to allow Web-based querying among collection databases managed by different institutions (Lowe 2004). These exchange protocols has given rise to Web-based, distributed information systems, such as Ornis (<http://www.specifysoftware.org/Informatics/informaticsortnis/>),

MaNIS (<http://www.manisnet.org/>), and HerpNet (<http://www.herpnet.org/>). The Global Biodiversity Information Facility (GBIF) is attempting to coordinate and support interoperability among worldwide biological collections (<http://www.gbif.org>). The intent of GBIF is to use the Internet to link distributed biological collections with non-biological databases (i.e. geospatial, ecological, climatic, and DNA sequence and other molecular data) to support data mining on a scale not yet possible (Edwards et al. 2000).

A significant challenge faced by information managers of biological collections is georeferencing accessions by assigning latitude and longitude coordinate values to collection sites (Beaman et al. 2004). Historic and incomplete site descriptions varying in accuracy and precision can make this process difficult. Considering the trend of serving information over widely distributed collection networks, standardized georeferencing protocols are absolutely essential to ensure success (Steiner and Greene 1996). Wicczorek et al., (2004) have developed a point-radius method for georeferencing. On line web resources and guidelines have been developed using this method, and are being widely used by many projects (i.e., <http://www.herpnet.org>, <http://www.elib.cs.berkeley.edu/manis/Documents.html>, and <http://www.gbif.org/prog/digit/1149073771>).

Research administrators are recognizing that resources are needed to support georeferencing efforts and are making funding sources available. In the United States since 1999, the National Science Foundation, Division of Biological Infrastructure (<http://www.nsf.gov/bio/dbi/start.htm>) has implemented the Biological Research Collections Program and awards \$6-million annually to support collection enhancement, computerization, and research aimed at developing better collection-management methods. In 2003, 11 of 24 funded projects focused on georeferencing accessions and providing Internet access to collections. GBIF also provides start-up money to improve electronic documentation including the georeferencing of biological collections (<http://www.gbif.org/prog/digit>).

With increasing numbers of collections being georeferenced and brought online, applications that map and spatially analyze these collections are now being developed. In 2004, GBIF launched a Web-based pilot project demonstrating four examples of how collection data can be mapped and incorporated into simple analyses. These can be used to compare the intensity of botanical collection with species distribution and allow incorporation of insect species-distribution maps into traditional land-use planning applications (<http://www.gbifdemo.utu.fi/index.htm>). Several Web-based applications have been developed that focus on predicting the geographic distribution of species. For example, Lifemapper (<http://www.lifemapper.org>) maps accessions from numerous natural history and herbarium collections and develops predictive distribution models based on accession and environmental data using the Genetic Algorithm for Rule-set Production (GARP) (Stockwell and Peters 1999). The WhyWhere software (http://www.biodi.sdsc.edu/ww_home.html) not only develops a fine-scale predictive map, but goes one step further to identify the most relevant predictors used in the model based on an extensive set of analyzed environmental variables (Stockwell 2004). Considering the complexities of species distributional modeling, and the fast-paced development of the field of conservation biogeography, these two programs represent fledgling efforts to capitalize on the Internet to deliver conservation relevant information (Whittaker et al. 2005).

For plant genetic resource (PGR) collections, web-based mapping and spatial analysis has additional value since the linkage between collection site environment

and an organism's environmental adaptation is well recognized (i.e., Vogel et al., 2005; Del Rio and Bamburg 2004; Hijmans et al., 2003). Mapping and spatial analysis of PGR collections enhance our understanding of how agricultural diversity and use are distributed over time and space (Guarino 1995), and also support more effective use of PGR collections by helping customers identify valuable germplasm that can be used to develop new cultivars and uses of crop species and support basic and applied botanical research (Guarino et al. 2002; Steiner 1999).

To date, there have been few applications developed that allow Web-based mapping of germplasm collections. The Germplasm Resources Information Network (GRIN), public web site (www.ars-grin.gov/npgs) uses ArcIMS (ESRI, Inc, Redlands, CA) to allow users to view a map highlighting an individual accession, along with all other accessions in the U.S. National Plant Germplasm System (NPGS). This feature is located on the Accession Information page that results from an Accession Query. The ArcIMS HTML Viewer allows simple applications to be developed quickly. However, more complex applications, such as those envisioned by the PGR community, will require more sophisticated and customized server-side and client-side scripting not possible with commercial prepackaged software. A second drawback of using commercial platforms is the high cost of implementing such systems. For example, in August of 2005, ESRI offered U.S. customers a Dell PowerEdge 1850 server loaded with ArcIMS Software for \$16,000 (http://www.esri.com/partners/hardware/hw_promo.html#dell). Finally, a third drawback is the need for purchasing a user-license for some commonly used applications that may restrict availability to under-capitalized users (Metz 2002).

Given the economic value of using open-source software, the purpose of this research was to: (i) demonstrate the potential of developing a Web-based mapping application for biological collections using WebGD, an open-source software development tool; and (ii) illustrate how simple spatial analysis can be used to help collection users describe the range of ecogeographic variation in biological collections and customize the selection of accessions based on georeferenced variables.

Materials and methods

Plant genetic resource collection

The United States Department of Agriculture (USDA), Agricultural Research Service (ARS), NPGS collections of *Medicago* (alfalfa) and *Trifolium* (clover) germplasm accessions were used. These collections contain more than 14,000 accessions, representing 79 species of *Medicago* and 229 species of *Trifolium*. The accessions are conserved as seeds and distributed worldwide to plant breeders and other scientists through the NPGS. Collection documentation has been recently upgraded and includes georeferences for all accessions that had adequate collection-site locality data (Greene 2001) from direct measurement by collection expedition members or retroclassification (Steiner and Greene 1996). This information can be retrieved over the Internet from the GRIN relational database. A sophisticated Oracle relational database was designed for GRIN to help curators and customers access information about more than 450,000 accessions in the USDA NPGS Germplasm Collection (Database Management Unit, USDA, ARS, NGRL 2004).

The GRIN data dictionary can be viewed at <http://www.ars-grin.gov/npgs/dict/prod/dd.html>.

Because of our desire to conduct interactive georeferenced analyses of plant genetic resources in the NPGS collection, a prototype, Web-based GIS application that uses the GRIN database schema was developed. Except for a few tables and fields added to support GIS features and analyses, our prototype emulates the existing structure of the GRIN database. GRIN data for the *Medicago* and *Trifolium* collections were downloaded and imported into an Access database (Microsoft, Redmond, WA) and an SQL Server database containing most GRIN database tables was created from the Access files. The content of the SQL Server database was then exported to a PostgreSQL database and used to develop the prototype Web-based application.

WebGD application development framework

We upgraded and repackaged the code developed by DM Solutions Group (<http://www.dmsolutions.ca>) as an application-development framework and refer to it as WebGD (Sano et al. 2003). The framework is application-independent and has currently been used in more than 10 applications, including the one featured in this paper (see <http://www.yukon.orst.edu>). With WebGD, software developers can create applications that allow users not only to retrieve data by using a simple Web-based interface, but also to select which geographical features are of interest (Wangmutitakul et al. 2003, 2004). We developed the Web interface with PHP, a widely used, open-source Web-scripting language. Although our application can query the data managed by PostGIS through MapServer, MapServer provides no functions for modifying those data. To rectify this problem, both geospatial and non-geospatial data are stored in the PostgreSQL object-relational database. Geospatial data are stored in geometry columns, and PostGIS uses these geometry columns to manage geospatial queries and updates. Non-spatial data are stored alongside the geospatial data. When the user clicks on a map feature, non-spatial data associated with the feature can be retrieved, updated, or deleted, depending on the request. The WebGD development framework will be a freeware product, jointly released by USDA-ARS and Oregon State University.

Architecture of WebGRMS prototype

The WebGD application-development framework was used to build our prototype application, the Web-Based Genetic Resources Management System (WebGRMS). The open-source architecture of WebGRMS is shown in Fig. 1. The geospatial data of WebGRMS are managed using PostgreSQL, an open-source object-relational database and PostGIS, an open source extension of PostgreSQL for GIS applications. MapServer (<http://www.mapserver.gis.umn.edu/>) developed at the University of Minnesota generates maps that are displayed on a Web browser by using geospatial data provided by PostGIS. Server-side scripts written in PHP generate Web pages, including the one that displays the maps. The PHP Mapscript module is used to interact with MapServer. The relationships among the WebGRMS components are as follows: (i) a Web browser requests a page from the Web server that may include geospatial information, (ii) the Web server forwards the request to the PHP script interpreter, (iii) if the request is for geospatial information, the PHP Mapscript

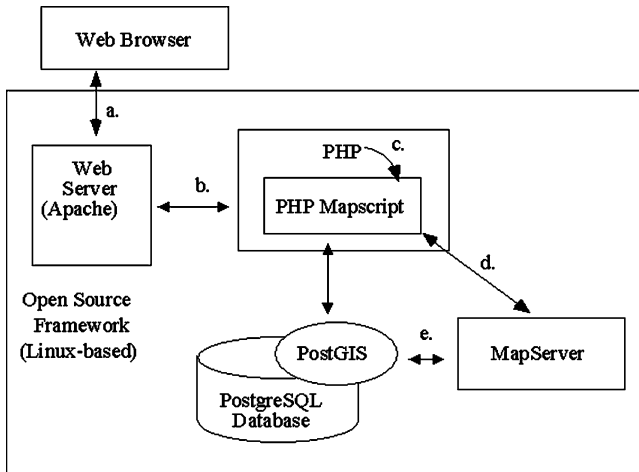


Fig. 1 Architecture of the prototype WebGRMS application; **a.** Web browser requests a page from the Web server; **b.** Web server forwards the request to the PHP script interpreter; **c.** PHP Mapscript module is called for geospatial data; **d.** PHP Mapscript forwards map request to MapServer; and **e.** MapServer generates a map based on geospatial data supplied by PostGIS

module is called, (iv) the PHP Mapscript forwards the proper map request to MapServer, and (v) MapServer generates a map that is based on geospatial data supplied by PostGIS. WebGRMS was developed and runs on a 1 GHz server with 120 Gbytes of hard drive storage. The cost of such hardware support is approximately \$800 (U.S.).

Results

The WebGRMS prototype can be viewed at http://www.yukon.een.orst.edu/~greene/ms_apps/home/index.htm. The prototype works best when using Microsoft Internet Explorer, (Version 5 or greater) as the web browser.

The demonstration site has three functional areas: (i) Query, (ii) Add Data, and (iii) Analyze. Each of the modules is a separate application developed by using WebGD tools.

Query module

The Query module allows users to visualize the geographic distribution of georeferenced accessions in the *Medicago* and *Trifolium* collections through an interactive mapping process. The initial map can be global scale or limited to a specific geographic region by selecting the geographic extent using the Quick View feature (Arur 2004), and redrawing the map. Various map layers can be selected including Olson's Ecoregion classification (Olson and Watts 1982), administrative boundaries, countries, cities, rivers, and collecting sites. Under Map Functions, radio buttons allow users to navigate freely by in-and-out zooming and panning. WebGD makes it easy to include maps of different resolution, which is illustrated in the prototype by

zooming in on the state of Oregon, or the Caucasus Mountains, Russia. For example, once one is completely zoomed in on Salem, Oregon, aerial photos are displayed (Fig. 2). There is also a radio button for Accession Info that provides accession-specific data when an individual accession site is clicked on. Several query features in Map Functions allow users to map subsets of the collections based on multiple criteria of the user's choice. In Search by Accession, an easy-to-use form allows search criteria to be defined based on identifier number, taxonomy, life form, improvement status, collector/donor, collection date, or country of origin. Users can also search for accessions that were collected in a specific Olson's Ecoregion Class, or in a user-defined range of elevation or precipitation values. In Search by Area, the same search criteria are available but limited to those accessions located in a user-defined geographic area. Once the query is processed, a list of accessions fitting the criteria is returned. Users can view detailed information on each accession brought

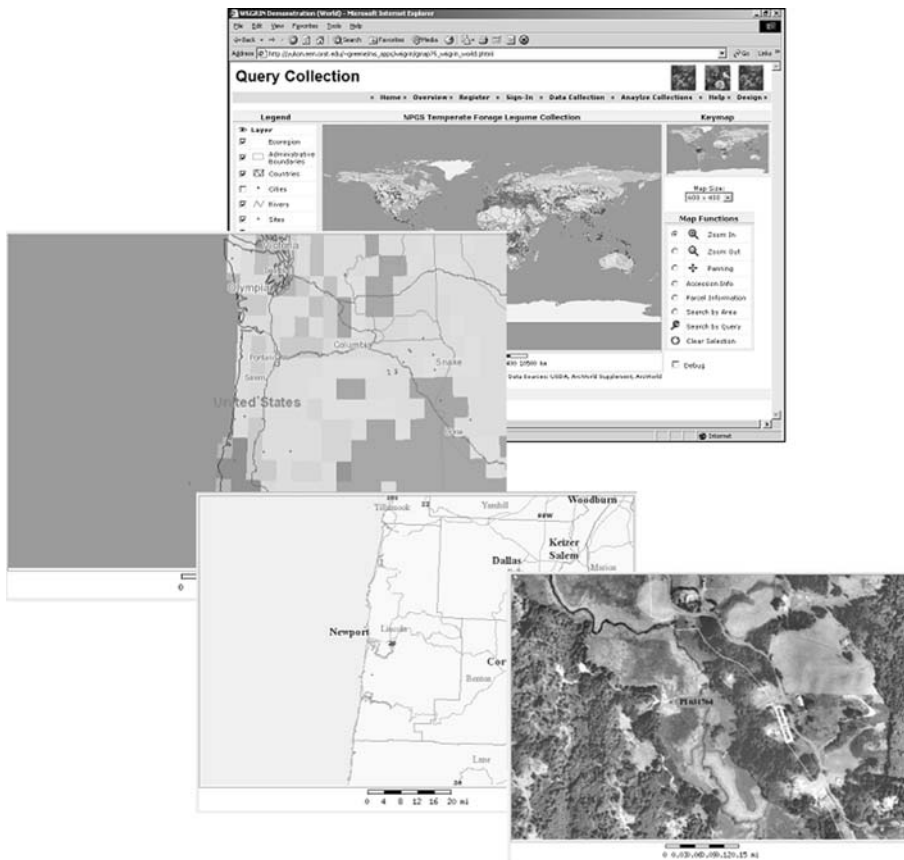


Fig. 2 WebGD tools allow inclusion of maps with varying resolution. As users zoom into an area, the map resolution increases accordingly. For example, zooming into Salem, Oregon, USA ultimately results in a digital orthographic aerial image. The prototype contains maps at this scale only for Oregon, and also has examples of high-resolution maps in the Caucasus Mountains, Russia

up with the query and select accessions of interest. Selected accessions can then be mapped by clicking Refresh Map.

The ability of a user to define and map subsets of accessions on an interactive basis is particularly valuable. Curators can use the feature to help visualize geographic gaps and redundancies among accessions. From a taxonomic perspective, these queries can answer the question, “How well do these collections represent the known geographic distribution of a given species?” To support acquisition efforts, users can map sites visited on previous collection trips to avoid redundant collecting and save valuable time and expense. Users can also readily identify accessions collected at the boundaries of a species’ range or make maps that show only landraces or wild material that have been collected in a specified area. Because of the interactive nature of the application, a user can repeatedly refine queries until a suitable subset is displayed.

Add data

In the current version of the WebGRMS prototype, forms for all the tables are generated automatically from the database schema by a WebGD form generator. A unique feature of our Web form generator is that it can generate forms that interact with the map interface as well as with the database. In particular:

1. The user can click on the map to specify the location of a collecting site whose record is to be inserted into the database.
2. The locations of sites retrieved by a database search are highlighted on the map and can be visualized more easily through the zoom feature.

Furthermore, the WebGD map interface is parameterized so that it can be easily customized for different user applications. Therefore, a complete set of Web forms for the tables in the database and a Web-based map interface can be created in a few months for a complex database application such as GRIN, if the relational database schema and the GIS data for the required map layers are available.

The prototype illustrates how data can be directly entered and/or updated into database tables using a Web-based interface. Data can be entered two ways. By using the radio buttons in the map functions box, a user can find the location of interest by zooming and panning on the map. After clicking the radio button for INSERT, the user then clicks on the location on the map. A form pops up, with the associated map coordinates, and site-specific data can be entered. The prototype has a form based on the GRIN Habitat table, since this would be a logical place to enter habitat data of a collection site. However, the form could include any table in a database. Once the form is completed, a click on INSERT adds the data to the table, and the new point can be viewed on the map. To check data entry, Search by Area can be used to query for the new record. The prototype also illustrates how data can be directly entered into forms. A click on ADD DATA USING FORMS allows a user to select the appropriate database table. A form pops up, and data can be entered. Of course, appropriate levels of security can be added to these features through the use of registration, password assignment and firewalls.

Analyze collections

The purpose of the Analyze Collection module is to illustrate how the results of simple analyses can be visualized with easily comprehensible graphs (Sharma 2004). These graphs assist users in identifying individual accessions and species that are most suited to a targeted environment and in assessing collections for representation of user-defined climatic factors. A data flow diagram illustrates the Analyze Collection module (Fig. 3). The various graphs generated provide information about annual and monthly temperature and precipitation for specific collection sites, and can also present general climatic data averaged across all accessions within a taxon (i.e., species ecological amplitude). Users can also specify the latitude and longitude of their choice by either entering map coordinates or selecting a location from a map. This information can be used to judge the suitability of individual accessions or species to a user-specified climate range and also help assess the general environments (that may not be geographically contiguous) represented by accessions in a collection.

Accession suitability

This analysis compares a specified location with an accession location for the annual monthly values of a single climate factor. The overall goal of this analysis is to help answer the question, “How does the climate of the collection site of this accession, compare to a user-defined point?” which should be useful for determining the introduction value of individual accessions to the targeted environment. A user specifies the accession of interest, climate factor, and the second location, and a line graph is generated for the desired attribute that plots monthly values for both locations (Fig. 4). The more two lines coincide, the more similar the environments.

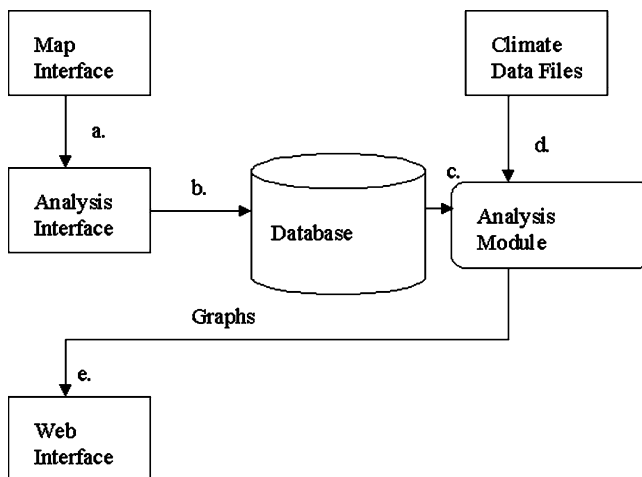


Fig. 3 Data flow diagram illustrates the analysis module. **a.** Map interface used to specify map coordinate data; **b.** Analysis interface used to set up query based on map coordinate data and database; **c.** **d.** Analysis module processes data from database and climate files to produce graphs; **e.** Graphs are viewed on Web interface

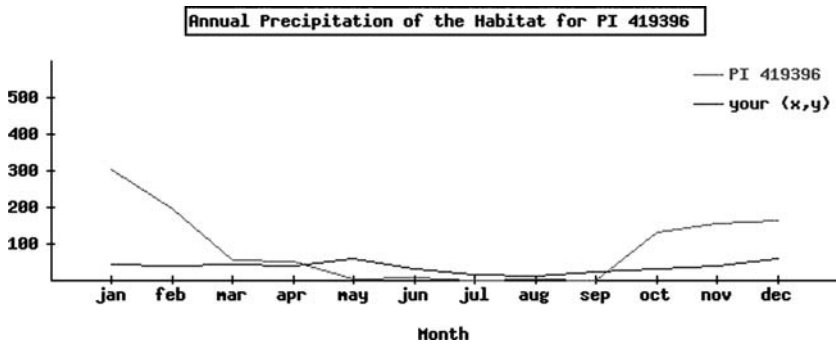


Fig. 4 The Analyze Collection module includes Accession Suitability, which graphs monthly annual precipitation or temperature of the collection site of a specified accession along with monthly precipitation or temperature for a user-specified location

Monthly climatic conditions for a species

The purpose of this bar graph is to display the distribution of monthly values of either temperature or precipitation based on collection site data for all accessions of a given species (Fig. 5). The user specifies the species, climate variable (temperature or precipitation), and a location of interest. The mean (μ) and standard deviation (σ) are then calculated from all georeferenced accessions of the user-defined species. Colors are assigned to the following ranges: $\mu - \sigma$ to $\mu - 2\sigma$ (orange); μ to $\mu - \sigma$ (yellow); $\mu + \sigma$ to μ (light blue); and $\mu + 2\sigma$ to $\mu + \sigma$ (dark blue) and a band graph displayed. If the user specifies a location for comparison, the corresponding monthly data are displayed as a line graph so the user can examine the band graph to determine the monthly climate range of a species and determine how the specified location compares. A limitation of this graph is that the ecological amplitude of the species is defined strictly by the accessions in the collection. If the collection has limited representation from its complete ecogeographic range, one would expect the ecological amplitude to reflect that bias. This problem illustrates the value of developing collection networks; as more collections come on line, this approach to modeling should become more robust, as the number of potential collection sites increases.

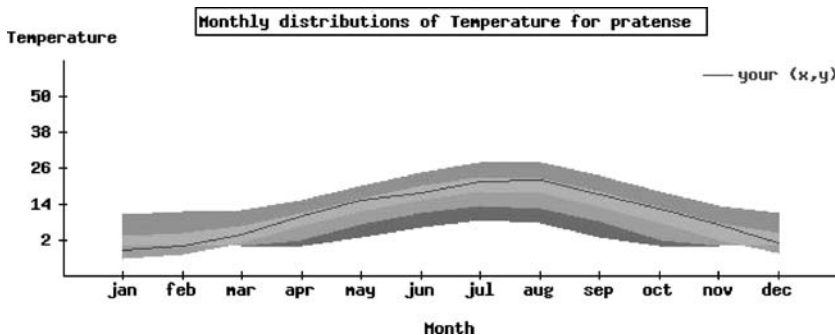


Fig. 5 The Analyze Collection module graphs monthly temperature of rainfall for all collection sites of a specified taxon

Collection representation of climatic distribution

Many taxa are distributed over a wide range of habitats. The following two graphs (Fig. 6 and 7) were designed to illustrate the extent that ex situ collections represents the climatic distribution of a specific taxon. These are useful for selecting accessions adapted to a target environment, as well as for assessing collections for redundancy and gaps in geographic or climatic representation.

The first graph displays a histogram representing the frequency distribution for all georeferenced accessions representing a specified taxon for a single climate factor (Fig. 6). Users specify the taxon, climatic variable, and seasonal range (by monthly or annual mean) and can also specify their own location. The resulting graph shows accession distribution for the specified climatic variable and highlights the bar of the histogram where the user-specified location falls to help users compare their location with accessions' collection sites. This analysis would be particularly useful to see if a biological collection held accessions that had been collected from a specific environment. For example, if cold temperatures limit growth in the target environment, a plant genetic resource user could examine the frequency distribution for January

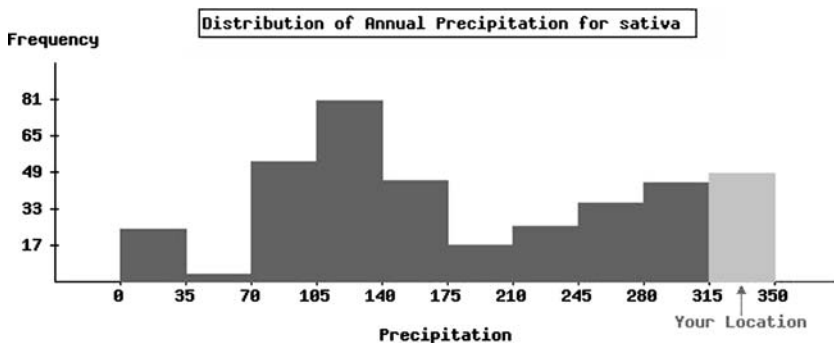


Fig. 6 The Analyze Collection module includes the following two graphs which were designed to illustrate the extent that a biological collection represents the climatic distribution of a specific taxon. Fig. 6 displays a histogram representing the frequency distribution for all georeferenced accessions representing a specified taxon for a single climate factor

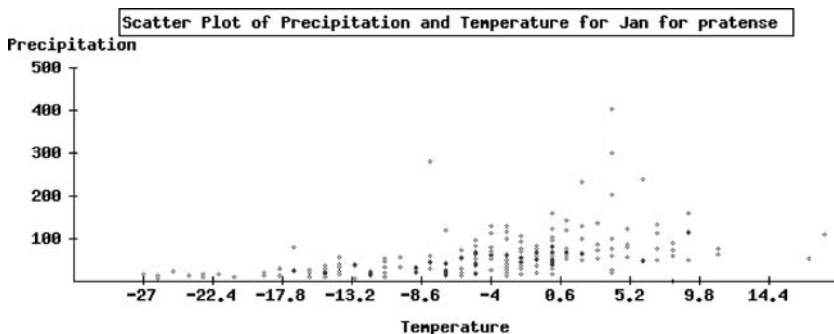


Fig. 7 Displays a scatter plot representing the distribution of two climatic factors (temperature and precipitation) for all georeferenced accessions. The user can specify the species and seasonal range (month or average annual) and can specify a location for comparison

mean temperatures at collection sites to determine if there are any accessions that come from environments resembling the target environment. Specifying the environmental parameters using the Search Form could then retrieve a relevant list of accessions. The Search Form could also be used to refine searches based on other parameters, such as improvement status. This graph would also be useful for assessing the proportions of accessions collected from similar environments. Evidence suggested for red clover (*Trifolium pratense* L.) that accessions collected from environmentally similar but geographically distant environments had certain morphological features in common (Greene et al. 2004). Curators, plant breeders, and other users would benefit from exploring these kinds of relationships to identify potentially useful germplasm and to assess collections for redundancy or gaps related to adaptation. This information could be used to set acquisition priorities for germplasm adapted to environments poorly represented in existing collections.

The second climate distribution graph (Fig. 7) displays a scatter plot representing the distribution of two climatic factors (temperature and precipitation) for all georeferenced accessions. The user specifies the species and seasonal range (month or average annual) and can specify their own location. The resulting scatter plot allows users to visualize how accessions in the collection are distributed relative to specified climatic variables. This analysis would be particularly useful to identify accessions that are adapted to climatic regimens, such as cold-dry-winters or hot-dry-summers.

Conclusion

WebGRMS demonstrates a Web-based application that can be used to facilitate the management and use of biological collections. WebGRMS provides an interactive mapping interface and has been used to present examples of simple spatial analyses and end-product visualization. Since most of the complex workings for delivering GIS functions over the Web had already been incorporated into the WebGD framework, it was possible to develop WebGRMS without extensive knowledge about the framework itself. Although WebGD was initially implemented with ArcIMS, ArcSDE, and SQL Server on a Windows 2000 Server, the WebGD framework on Windows was less reliable, slower and more complex than the current version (Wuttiwat et al. 2003). By using the current version of WebGD from the initial conception of this project in May 2003, a mock up of the application took less than eight weeks for a team of graduate students to develop with approximately 800 h of programming effort. Since that time, approximately 500 h of programming effort have been focused on generating the automated form generator and refining the prototype. Since WebGRMS was based on open-source code, the cost of implementing the system was greatly reduced. Not only is open-source code freely available on the Internet, but user groups provide a valuable and free source of support.

WebGRMS illustrates the many advantages of a Web-based information system that would be suitable for under-resourced biological collections. Metz (2002) discussed the advantages of using an open-source approach in developing genetic resource documentation systems, noting that the approach is intrinsically attractive to small germplasm collections and germplasm institutes in developing

countries due to low cost and ease of development. The approximate cost of server hardware to support a system capable of implementing the present GRIN database schema is approximately \$800 (US), including two backup hard disks. The application can be maintained at a single server site, and as updates in coding become effective, all users immediately benefit, thus simplifying system maintenance.

Collection users also benefit from an interactive, Web-based environment. Users can access a complex application and large amounts of data with only a standard Web-browser. No software or data need to be installed and no costly software use licenses purchased. Data can be entered into the system by a geographically dispersed group of users. Because there is a single server, security precautions and data validation must be built into the system. However, once the data are reviewed, the information can become immediately accessible to all users. Since the WebGRMS prototype was created, Minoura's group has continued to build upon the concept by developing several additional web-based GIS/database (WebGD) applications that allow users to insert, query, update, and delete geographical features and the data associated with them from standard Web browsers (<http://www.web.engr.oregon-state.edu/~minoura/research/>). The group has also built a WebGD application generator (WebGD-Gen) that automatically produces a WebGD application including a map interface from a database schema. This application generator greatly simplifies the process of creating a complex Web-based GIS/database application and significantly reduces the development time and maintenance cost. The WebGD framework and WebGD-GEN now support such features as tight integration of a Web-based map interface with a database, automatic selection of the spatial reference and map layers for the current region, and automatic generation of Web forms.

The current emphasis on developing Web-based, biodiversity information systems that rely on distributed databases suggests that further development of WebGRMS needs to include the capacity for handling data from multiple databases (Graham et al. 2004; Lowe 2004). Additionally, WebGRMS could be enhanced by utilizing a larger array of eco-geographic data and developing more sophisticated spatial analyses. In its current form, the prototype WebGRMS application provides users interested in *Medicago* and *Trifolium* germplasm with an innovative method to examine NPGS germplasm collections. More importantly, we hope the prototype provides a glimpse into the future of Web-based GIS analysis of biological collections, and provides incentives for addressing current shortcomings in the quality and completeness of accession georeferencing and in data standardization and database interoperability. The potential value of mining knowledge from biological collections rests upon accepting, understanding, and addressing these challenges in innovative ways to ensure that past work to build these collections and current investments in conserving them are not wasted.

Acknowledgments This research is supported in part by Web-GD tools developed from a grant in cooperation with the Yolo County Resource Conservation District through the CALFED Bay-Delta Restoration Project, Ecosystems Restoration Program (Contract #01-N25). Thanks to D. Arur, J. Sano, and A. Sharma who made significant programming contributions to the development of code used in the WebGRMS application. Thanks also to R. Hannan, A. Stoner, and J. Mowder for initial review of the concepts developed in WebGRMS and suggestions related to its implementation.

References

- Arur DV (2004) Dynamic map reprojection for Web-based GIS/Database (WebGD) applications. Unpublished thesis, Computer Science Department, Oregon State University, Corvallis
- Beaman R, Wiczorek J, Blum S (2004) Determining space from place for natural history collections. *D-lib Magazine* 10(5) [Online]. Available at <http://www.dlib.org/dlib/may04/beaman/05beaman.html> (Verified January 3, 2005)
- Bisby FA (2000) The quiet revolution: biodiversity informatics and the Internet. *Science* 289:2309–2312
- CREEDA Group. 2003. CREEDA web site [Online]. Available at <http://www.web.engr.oregon-state.edu/~minoura/research/creeda/index.html> (Verified August 31, 2004.) Oregon State University, Corvallis, OR
- Database Management Unit, USDA, ARS National Germplasm Resources Laboratory. 2004. National Plant Germplasm System, GRIN-USDA-ARS [Online]. Available at <http://www.ars-grin.gov/npgs/> (Verified August 31, 2004), USDA, ARS, Beltsville, MD
- Del Rio AH, Bamberg JB (2004) Geographical parameters and proximity to related species predict genetic variation in the inbred potato species *Solanum verrucosum* Schlecht. *Crop Sci* 44:1170–1177
- Edwards JL, Lane MA, Nielsen ES (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289:2312–2314
- Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trend Ecol Evol* 19:497–503
- Greene SL (2001) Improving the quality of germplasm data to enhance germplasm use and management. *Plant Genet Res Newsl* 125:1–8
- Greene SL, Gritsenko M, Vandemark G (2004) Relating morphologic and RAPD marker variation to collection site environment in wild populations of red clover (*Trifolium pratense* L.). *Genet Res Crop Evol* 51:643–653
- Greene SL, Morris JB (2001) The case for multiple-use plant germplasm collections and a strategy for implementation. *Crop Sci* 41:886–892
- Guarino L (1995) Mapping the ecogeographic distribution of biodiversity. In: Guarino L, Rao V, Reid R (eds) *Collecting plant genetic resources*. CAB International, Wallingford UK, pp 287–314
- Guarino L, Jarvis A, Hijmans RJ, Maxted N (2002) Geographic information systems and the conservation and use of plant genetic resources. In: Engels JM et al (eds) *Managing plant genetic diversity*. CABI Publishing, Oxon UK, pp 387–404
- Guerrant EO, Havens K, Maunder M, Raven P (2004) Ex situ plant conservation: supporting species survival in the wild (society for ecological restoration international). Island Press, Washington US
- Hijmans RJ, Jacobs M, Bamberg JB, Spooner DM (2003) Frost tolerance in wild potato species: assessing the predictivity of taxonomic, geographic, and ecological factors. *Euphytica* 130:47–59
- Krishtalka L, Humphrey PS (2000) Can natural history museums capture the future? *Bioscience* 50:611–617
- Lowe JW (2004) Bone rooms, bird bodies and biodiversity informatics. *GeoSpatial Solutions* [Online] Available at <http://www.geospatial-online.com> (Verified August 31, 2004), Advanstar Communications
- Metz T (2002) Open source software as an alternative to set up plant genetic resources documentation systems. *IPGRI Newsl* 8(1):6–7
- Olson JS, Watts JS (1982) Major world ecosystem complexes. In: *Carbon in live vegetation of major world ecosystems*. ORNL-5862. Oak Ridge, TN: Oak Ridge National Laboratory. 1:30,000,000
- Ponder WF, Carter GA, Flemons P, Chapman RR (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conserv Biol* 15:648–657
- Sano J, Wanalertlak N, Maki A, Minoura T (2003) Benefits of web-based GIS/database applications. URISA 2nd Annual Public Participation GIS Conference, Portland State University, July 20–23, Portland, OR
- Sharma A (2004) Web-based analysis module for a germplasm collection. MS Thesis, Oregon State University, Corvallis, OR
- Steiner JJ (1999) Exploring the relationship of plant genotype and phenotype to ecogeography. In: Greene SL, Guarino L (eds) *Linking genetics and geography: emerging strategies for managing germplasm diversity*. CSSA Special Publ. 27, Madison WI, pp 39–50
- Steiner JJ, Greene SL (1996) Proposed use of ecological descriptors and their utility for plant germplasm collections. *Crop Sci* 36:439–451

- Stockwell DRB (2004) Biodiversity Insight System: Whywhere [Online]. Available at http://www.biodi.sdsc.edu/ww_home.html (accessed 20 August, 2004; Verified 31 August, 2004), San Diego Supercomputing Center, University of California, San Diego
- Stockwell DRB, Peters D (1999) The GARP Modeling System: problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 13(2):143–158
- Vogel KP, Schmer MR, Mitchell RB (2005) Plant adaptation regions: ecological and climatic classification of plant materials. *Rangeland Ecol Manage* 58:315–319
- Wangmutitakul P, Li L, Minoura T (2003) User-participatory web-based GIS/database applications. In 2003 GeoTec Event Proceedings: Global Conference and Exhibition on Geospatial Technology, Tools and Solutions. March 16–19, Vancouver, British Columbia
- Wangmutitakul P, Minoura T, Maki A (2004) WebGD: a framework for web-based GIS/Database applications. *J Object Technol* 3(4):209–226
- Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation Biogeography: assessment and prospect. *Divers Distrib* 11:3–23
- Wieczorek J, Guo Q, Hijmans RJ (2004) The point radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int J Geogr Inf Sci* 18:745–767
- Wuttiwat T, Minoura T, Steiner J (2003) Using digital orthographic aerial images as user interfaces. *Proc. Annual Meeting Amer. Soc. Photogrammetry & Remote Sensing*. 5–9 May. Anchorage, AK